地方議会会議録コーパスの構築および政治情報システム構築を 目標としたアノテーションの一提案

筒井 貴士[†]・我満 拓弥^{††}・大城 卓[†]・菅原 晃平[†]・永井 隆広[†]・
渋谷 英潔^{†††}・木村 泰知^{††††}・森 辰則^{†††}

近年、国会や地方議会などの会議録が Web 上に公開されている。会議録は、首長や議員の議論が書き起こされた話し言葉のデータであり、長い年月の議論が記録された通時的なデータであることから、政治学、経済学、言語学、情報工学等の様々な分野において研究の対象とされている。国会会議録を利用した研究は会議録の整備が進んでいることから、多くの分野で行われている。その一方で、地方議会会議録を利用した研究については、各分野で研究が行われているものの、自治体により Web 上で公開されている形式が異なることが多いため、収集作業や整形作業に労力がかかっている。また、各研究者が重複するデータの電子化作業を個別に行っているといった非効率な状況も招いている。このような背景から、我々は多くの研究者が利用することを目的として、地方議会会議録を収集し、地方議会会議録コーパスを構築した。本稿では、我々が構築した地方議会会議録コーパスについて論ずる。同コーパスは、Web 上で公開されている全国の地方議会会議録を対象として、「いつ」「どの会議で」「どの議員が」「何を発言したのか」などの各種情報を付与し、検索可能な形式で収録した。また、我々は会議録における発言を基に利用者と政治的に近い考えをもつ議員を判断して提示するシステムを最終的な目的としており、その開発に向けて、分析、評価用のデータ作成のために会議録中の議員の政治的課題に対する賛否とその積極性に関する注釈付けをコーパスの一部に行なった。本稿では、注釈付けを行なった結果についても報告する。

キーワード：注釈付け、地方議会会議録コーパス、地方政治、政治情報システム

Construction of Regional Assembly Minutes Corpus and a Proposal for Annotation Scheme for Implementing Political Information System

TAKASHI TSUTSUI[†], TAKUYA GAMAN^{††}, TAKASHI OSHIRO[†], KOHEI SUGAWARA[†],
TAKAHIRO NAGAI[†], HIDEYUKI SHIBUKI^{†††}, YASUTOMO KIMURA^{††††} and TATSUNORI MORI^{†††}

[†] 横浜国立大学大学院環境情報学府, Graduate School of Environment and Information Sciences, Yokohama National University

^{††} 東京大学大学院工学系研究科, Graduate School of Engineering, The University of Tokyo

^{†††} 横浜国立大学大学院環境情報研究院, Graduate School of Environment and Information Sciences, Yokohama National University

^{††††} 小樽商科大学商学部, Faculty of Commerce, Otaru University of Commerce

In recent years, minutes of regional assemblies and the National Diet have been published on the web. Those minutes have long recorded transcribed discussions of mayors and members of assemblies. Therefore, they are a target of study in various fields such as politics, economics, linguistics, information engineering. Since the minutes of the National Diet are maintained in electronic form and freely available via a search system, many researchers have utilized the minutes as a target of study. Minutes of regional assembly meetings are also the focus of researchers in various fields. However, researchers have had trouble gathering and preparing minutes for their study, because the way in which minutes are made available to the public varies assembly by assembly. It is very inefficient for each researcher to make the effort to digitize minutes separately. To improve the situation and contribute to research communities, we have collected regional minutes of assemblies and constructed the corpus of regional assembly minutes. In this paper, we discussed the construction of the corpus of regional assembly minutes. The corpus records minutes from regional assemblies all over Japan that are available on the web. We added additional information to the corpus, such as “date,” “name of meeting,” “name of speaker,” “text of statement,” so that users may search statements across the corpus using such information. The final goal of our project is to build a political information system that can recommend a suitable person, or members of an assembly, according to the consistency between users’ opinions and statements of assembly members. As a preliminary step of development, we annotated a part of the corpus with information about the speaker’s attitude to specific political subjects, including degree of approval/disapproval. In this paper, we also report the result of the annotation.

Key Words: *Annotation, Regional Assembly Minutes Corpus, Local Politics, Political Information System*

1 はじめに

平成 11 年から政府主導で行われた平成の大合併や、平成 19 年より施行された地方分権改革推進法など、地方政治を重視する取り組みが盛んに行われていたのは記憶に新しい。一方で、有権者の政治離れが深刻な問題となって久しく、平成 25 年 7 月 21 日の第 23 回参議院議員通常選挙における選挙区選挙では 52.61%の投票率¹となり、参議院議員通常選挙において過去 3 番目に低い値となった。地方政治の場合、平成 23 年 4 月の第 17 回統一地方選挙の投票率は、48.15%²であり、さらに低い値となっている。

地方政治に対する有権者の政治離れの原因には幾つか考えられるが、その一因に地方議会議員およびその活動の認知度の低さがあげられる。現状では、政治情報を入手するソースとして

¹ <http://www.soumu.go.jp/senkyo/senkyo.s/data/sangiin23/index.html>

² <http://www.soumu.go.jp/senkyo/senkyo.s/data/chihou/ichiran.html>

テレビや新聞などのマスメディアが占める割合が大きいですが、このようなマスメディアに首長以外の地方議会議員が取り上げられることはほとんどない。地方議会議員は国会議員と同様に住民による選挙によって選ばれ、かつ、国政よりも身近な存在であるべきであるにもかかわらず、その活動に関する認知度が低いのは大きな問題であると考えられる。そこで、住民に提供される地方政治の情報、特に地方議会議員に関する情報量の不足を解決するための方法の一つとして、Web上の情報を有効に利用することを考える。Web上に存在する議員の情報には、議員や政党のホームページ、ニュースサイトの政治ニュース、議員のブログやTwitterなどのSNS、manifesto、議会の会議録などがある。このうち会議録には、議員からの一方的な情報発信ではなく、議論や反対意見などのやりとりが含まれ、公の場における各議員の活動や考え方を知ることができる。また、研究対象として会議録を見た場合、会議録は、首長や議員の議論が書き起こされた話し言葉のデータであり、長い年月の議論が記録された通時的なデータであることから、政治学、経済学、言語学、情報工学等の様々な分野における研究対象のデータとして利用されている。

例えば、政治学の分野では、平成の大合併前後に行われた市長選挙についての分析を行い、合併を行った市と行わなかった市の違いを当選者の属性から比較した平野(平野 2008)の研究、合併が地方議会や議員の活動に対して与えた影響を 856 議員にアンケート調査することで分析を行った森脇(森脇 2008)の研究などがある。また、経済学の分野では、「小規模自治体の多選首長は合併に消極的」という仮説を検証するために、全国の地方議員、首長の情報を人手で調査した川浦(川浦 2009; Kawaura 2010)の研究など、言語学の分野では、「去った〇日」という表現(「去る〇日」の意)が那覇市の会議録に見られることを指摘した井上(井上 2013)、「めっちゃんこ」が名古屋市の会議録に見られることを指摘した山下³、形態素 N-gram を用いて地方議会会議録の地域差を捉える方法について検討した高丸ら(高丸, 木村 2010; 高丸 2011, 2013)、発言者の出身地域とオノマトペの使用頻度についての分析を行った平田ら(平田, 中村, 小松, 秋田 2012)などの研究が存在する。情報工学の分野においても、特徴的な表層表現を手掛かりに国会会議録を対象とした自動要約を行った川端ら(川端, 山本 2007)や山本ら(山本, 安達 2005)の研究、住民の潜在的な関心を明確化するための能動的質問生成手法を提案した木村ら(木村, 渋谷, 高丸, 乙武, 小林, 森 2011)の研究などが存在し、海外でも、会議録中の発言を元にイデオロギーを分類する Yu et al.(Yu, Kaufmann, and Diermeier 2008)や、会議録で用いられている語句を可視化する Geode et al.(de Goede, van Wees, Marx, and Reinanda 2013)などの研究が行われている。

これらの研究を行う上で基礎となる会議録のデータであるが、国会の場合、国立国会図書館

³ <http://dictionary.sanseido-publ.co.jp/wp/2012/07/07/>

により会議録サイト⁴が整備されており、第1回国会（昭和22年）以降のすべての会議録がテキストデータとして公開され、検索システムによって検索を行うことができる。一方で、地方議会会議録の場合、全ての自治体の会議録をまとめているサイトは存在せず、自治体ごとに参照する必要がある。加えて、自治体により Web 上で公開されている形式が異なることが多いため、統一的に各自治体の会議録を扱おうとすれば収集作業や整形作業に労力がかかる。また、各研究者が重複するデータの電子化作業を個別に行っているといった非効率な状況も招いている。

このような背景から、我々は地方政治に関する研究の活性化・学際的応用を目指して、研究者が利用可能な地方議会会議録コーパスの構築を行っている。コーパスの構築にあたっては、木村ら（木村，渋谷，高丸 2009）や乙武ら（乙武，高丸，渋谷，木村，荒木 2009）において行われた、北海道の地方議会会議録データの自動収集や加工の技術を参考にし、全国の市町村の議会会議録を対象としたコーパス構築を行うこととした。地方議会会議録コーパスは、Web 上で公開されている全国の地方議会会議録を対象として、「いつ」「どの会議で」「どの議員が」「何を発言したのか」を、発言に対して市町村や議会種別、年度や発言者名などの各種情報を付与することで構築し、検索可能な形式で収録する。また、近年、ヨーロッパでは VoteMatch⁵と呼ばれる投票支援ツールが多くの利用者を獲得しており（上神 2006; 上神，堤 2008; 上神，佐藤 2009），日本でも「投票びったん⁶」などの日本語版ボートマッチシステムが利用されていること、さらに平成 25 年 4 月 19 日から公職選挙法が一部改正され⁷，インターネットなどを利用した選挙運動のうち一定のものが解禁されたことなどから、我々は、地方議会会議録コーパスを用いて、会議録における発言を基に利用者と政治的に近い考えをもつ議員を判断して提示するシステムを最終的な目的としている。

さて、地方議会会議録コーパスを構築すると、会議録を文字列や単語で検索することができるようになる。さらに、会議録の書誌情報や議員情報に基づいて簡単な注釈付けを行うことにより、年度や地域をまたいだ比較検討や、地域ごとの表現の差の分析などを行うことが可能となる。その一方で、我々が構築を目指しているシステムは利用者と政治的に近い議員を判断し利用者に提示するものであるため、会議録の書誌情報や発言議員名といった簡単な注釈付けのみでは議員の施策や事業に対する意見の判別を行うことができず不十分である。すなわち、議員の発言の中にある施策や事業に対する意見のように、下位構造が存在し、それらが結びつくことで一つの情報となるものに対しての分析を行うことは、会議録の文字列検索のみでは難しい。

政治的な考えの近さは、一般に、施策や事業などへの賛否の一致度合いにより推測できると考えられ、上神ら（上神 2006; 上神，堤 2008; 上神，佐藤 2009）などのボートマッチシステムで

⁴ <http://kokkai.ndl.go.jp/>

⁵ <http://www.votematch.net>

⁶ <http://www.votematch.jpn.org/>

⁷ http://http://www.soumu.go.jp/senkyo/senkyo_s/naruhodo/naruhodo10.html

も、この考え方に基づいている。さらに、議員の施策や事業に関する賛否の意見には、同じ賛成の立場をとる議員の間でもその賛成の度合いには差が存在している。例えば、「昨年度は〇〇などの事業に取り組んできた」と発言した議員と、「〇〇などの事業を行うのもやむを得ない」と発言した議員では、前者の方が既に自らが取り組んでいることを表明していることから、より積極的に賛成であると考えられる。積極的に賛成である議員の方が、消極的に賛成である議員よりも、彼らが賛成する施策や事業の実現に向けて尽力すると考えられるため、当該の施策や事業を実現してほしい利用者には、積極的に賛成である議員の方を提示することが望ましい。また、消極的に反対の意見を示している議員よりも、積極的に反対の意見を示している議員の方が、彼らが反対する施策や事業を廃止することに注力すると考えられるため、反対の立場を取る議員に対しても同様の考えが成り立つ。このように賛否に加えて積極性を考慮して、利用者と近い考えをもつ議員を判断する必要がある。以上の背景から、我々は、比較的簡単な処理により自動的に付与できるタグを地方議会会議録コーパス全体に付与するとともに、上記の政治情報システムの検討のために、会議録の一部に対して、議員の施策・事業に対する賛否とその積極性に関連する情報の注釈付けを行うこととした。

本稿ではまず2節で関連研究について述べ、3節では地方議会会議録の収集及び地方議会会議録コーパスの構築について説明する。次に、4節では地方議会会議録コーパスの一部に対して我々が付与したタグの仕様や注釈結果の統計とその分析及び残された課題について述べる。最後に5節でまとめる。

2 関連研究

本節では、本稿に関連する各種研究について説明する。

2.1 会議録を対象とした研究

会議録を対象とした研究としては、以前より国会会議録を対象とした研究が行われてきた。川端ら(川端, 山本 2007)や山本ら(山本, 安達 2005)は特徴的な表層表現を手掛かりに国会会議録を対象とした自動要約を行っている。平田ら(平田 他 2012)は、発言者の出身地域とオノマトペの使用頻度についての分析を行っている。また、国会会議録検索システムというシステムが公開されており、国会の会議録を自由に検索・閲覧することができる。これに対し、地方議会会議録のもつ会議録検索システムは、市町村ごとに様式が異なっているため、複数の市町村の会議録を対象に研究を行おうとした場合にそのまま利用することは難しい。そこで、地方議会会議録を収集して統一された書式に整形する必要がある。

これに関連し、木村ら(木村 他 2009)や乙武ら(乙武 他 2009)は北海道内の各市町村を対象に地方議会会議録の自動収集に向けた公開パタンの分析を行っている。51種類の収集パターンに

よる自動収集プログラムを用いて約 94%の自治体から会議録の収集に成功している。この成果を参考にしつつ、我々は、各自治体が会議録を公開している形式を分析し、全国規模の会議録の収集を行った。

2.2 コーパス構築に関する研究

Web 文書を対象としコーパスを構築する研究では、以下の研究が存在する。関口ら (関口, 山本 2003) は Web 文書を収集し、HTML タグや日本語文章の書法を用い、質の面での改善を行うことで Web コーパスを作成した。橋本ら (橋本, 黒橋, 河原, 新里, 永田 2011) は、ブログを対象とした自然言語処理の高精度化への寄与を目的とし、81 名の大学生に 4 つのテーマで執筆させた 249 記事のブログに、文境界、形態素、係り受け、格・省略・照応、固有表現、評価表現に関する注釈付けを行った。Ptaszynski et al.(Ptaszynski, Rzepka, Araki, and Momouchi 2012) は、日本語のブログを自動収集して構築した、3.5 億文からなるコーパス YACIS に対して自動的に感情情報を付与した。また、飯田ら (飯田, 小町, 乾, 松本 2008) は新聞記事を対象とし、述語項構造・共参照タグを付与する基準について報告し、事態性名詞のタグ付与において、具体物のタグ付与と項のタグ付与を独立に行うことで作業品質を向上させている。しかしながら、本研究でコーパス構築の対象としたデータは地方議会会議録であり、これらのコーパス構築の手法とは対象とするデータが異なる。

2.3 主観的な情報の注釈付けに関する研究

本研究は政治的課題に対する賛否と積極性に関する注釈付けを行っており、主観的な注釈付けの一つである。主観的な注釈付けとしては、以下の研究がある。Weibe et al.(Weibe, Wilson, and Cardie 2005) は、意見などの private state をニュース記事の句に対して注釈付けを行っている。松吉ら (松吉, 江口, 佐尾, 村上, 乾, 松本 2010) は、書き手が表明する真偽判断、価値判断などの事象に対する総合的な情報を表すタグの体系を提案し、これに基づくコーパスを基礎とした解析システムを提案した。また、評判情報に関する研究では、小林ら (小林, 乾, 松本 2006) は主観的評価の構成要素を「根拠」「評価」「態度」の 3 つの要素に分類したうえでの注釈付きコーパスの作成を行っている。宮崎ら (宮崎, 森 2010) は、Web 文書を対象に、製品の状態と評価とを分離した評判情報のモデルを提案し、評判情報コーパス構築の際の注釈者間の注釈揺れを削減する方法を論じている。大城ら (大城, 渡邊, 渋谷, 木村, 森 2012) は、施策や事業に対する賛否の意見を、構造的に捉えるための注釈付けタグセットを提案し、その有効性を確認した。

我々の提案する注釈付けは、意見や評判情報の注釈付けと同様に文中のある部分に対して極性を付与するという点で共通しているが、極性に加えて程度を表す積極性の情報を注釈付けしている点でこれらの研究と異なる。積極性の情報を注釈付けすることの有用性については、次

節で説明する.

2.4 ボートマッチに関する研究

ボートマッチは選挙に関するインターネットサービスの一種で, 有権者と立候補者, または有権者と政党の考え方の一致度を測定することができるシステムである. 上神ら (上神 2006; 上神, 堤 2008; 上神, 佐藤 2009) はコンピュータによりコーディングを自動化する手法を提案しマニフェストの分析の自動化を行い, それを用いてボートマッチシステム「投票ぴったん」を作成した. また, 毎日新聞の「えらぼーと」⁸ などが公開されている.

木村ら (木村 他 2011) は意思決定の際に用いられる決定木を用い, 「決定木において同じ経路を選択する相手は同じ考え方をする相手とみなすことができる」という仮説のもとに, 利用者の政治的興味や関心を同定するための質問生成手法を提案している.

我々の場合, 賛否に加え積極性についても考慮し注釈付けを行うため, 既存のボートマッチシステムでは比較を行うのが難しいある施策や事業に対し同意見の議員を積極性という尺度を用いて分類することが可能となる. それにより, 「昨年度は〇〇などの事業に取り組んできた」と発言した議員と, 「〇〇などの事業を行うのもやむを得ない」と発言した議員のように, どちらも賛成の意思を示しているが積極性の度合いが異なる場合に, 我々の提案する注釈付け手法を用いれば, 前者の議員がより積極的に賛成であると注釈付けることが可能である. これにより, 当該の施策や事業を実現してほしい利用者に対し, その意見により近い前者の議員を提示することが可能となる.

3 地方議会会議録コーパス

本節では, 地方議会会議録の収集および地方議会会議録コーパスを構築するプロジェクトの概要及び, 構築された地方議会会議録コーパスについて説明する.

3.1 プロジェクトの目的

本プロジェクトは, 地方政治に関する研究の活性化・学際的応用を目指して, 研究者が利用可能な地方議会会議録コーパスを全国規模で構築し Web 上で提供することを目的とする. また, そのコーパスを利用した政治学, 社会言語学, 情報工学の研究を行い, その成果を学際的に応用した政治情報システムの開発を行う. プロジェクトの全体像を図 1 に示す.

⁸ <http://vote.mainichi.jp/>

3.2 地方議会会議録の収集

全都道府県の県庁所在地と政令指定都市の計 51 市町村の会議録について平成 17 年から平成 22 年を対象に収集を行った。市町村の会議録の多くは、Web 上で専用の会議録検索システムを通して公開されている。その会議録システムは表 1 に示すように大きく分けて 4 つの会社が会議録検索システムを提供しており、それぞれ付録 A に示すクロールプログラムを構築し収集を行った。なお、その他に該当するのは秋田市のみで、独自の会議録検索システムを作っていたため人手により会議録を収集した。

図 1 プロジェクトの全体像

表 1 51 市町村の会議録検索システム

会社名	市町村数
大和速記情報センター	14
会議録研究所	23
フューチャーイン	10
神戸総合速記	3
その他	1

3.3 地方議会会議録コーパスの構築

利用者の利便性を考慮し、付録 A の方法により収集した会議録に対し表 2 に示す付随情報を付与し、データベース化を行った。その際には必要な発言のみを簡単に参照できるように会議録を発言単位に分割した。発言単位の分割については、句点や括弧などを区切りにしており、その際に HTML タグはすべて取り除いている。以下、発言に付与する項目について説明する。「発言 ID」は各発言の識別を行うため、「市町村コード」は市町村ごとの検索のため、「議会名」は議会ごとの検索のため、「議会種別コード」市町村によって名称の違う議会名を分類するためにそれぞれ必要となる。「年度」、「回」、「月」、「号」、「日付」については時間情報として重要なため必要である。「表題」はページのタイトルとして、「段落番号」は段落ごとの抽出を容易にするため、「役職名」は会議によって議員の役職が変わることがあるため、「発言者名」は会議録中の文字列をそのまま保持するため、また、発言者が議員であるとは限らないため、「議員 ID」は議員の識別のためにそれぞれ必要である。「ファイルのパス」は元ファイルを参照することを容易にするため、「その他」は発言とそれ以外の内容を区別するためにそれぞれ必要となる。

例えば、図 2 のような会議録が与えられたとき、下線部の発言に対して表 3 のように情報が付与され、図 3 の様になる。

「発言 ID」は会議録では 2 番目の発言であることを表している。「市町村コード」は総務省により割り当てられた地方公共団体コードを指す。「議会種別コード」は定例会と臨時会には個別

表 2 発言に付与する項目

項目名	型	説明
発言 ID	int	自動採番
市町村コード	varchar	総務省により割り当てられた地方公共団体コード
議会種別コード	varchar	定例会 0010, 臨時会 0020, その他 1000
年度	int	西暦
回	int	開催数
月	int	開催月
議会名	varchar	例：定例会, 定例市会, 予算委員会
号	int	会議が何日目か
日付	varchar	開催日
表題	mediumtext	ページのタイトル
段落番号	int	発言の段落番号
役職名	varchar	議員の役職
発言者名	varchar	議員の役職名
議員 ID	int	あらかじめ議員に割り当てられた番号
ファイルのパス	varchar	元ファイルの保存場所
発言	mediumtext	1 文
その他	mediumtext	会議録内の発言以外の内容

○議長（櫻井正富君） 次に、日程第3 議案第1号から日程第55 議案第53号までの53案を一括議題といたします。

市長から、上程議案前部に対する提案理由の説明を求めます。

なお、提案理由の説明に合わせて、新年度に臨む所信の表明を行いたい旨の申し出がありますので、これを許します。

市長 宮島雅展君

（市長 宮島雅展君 登壇）

○市長（宮島雅展君） 本日ここに、3月市議会定例会を開会するに当たり、私の市政運営に対する所信の一端と、議案第1号から議案第15号までの平成22年度予算の概要につきまして述べさせていただきますと存じます。

図 2 甲府市議会会議録の例

表 3 発言に対する情報付与の例

項目	値
発言 ID	2
市町村コード	19201
議会種別コード	0010
年度	2010
回	1
月	3
議会名	定例会
号	1
日付	0301
表題	平成 22 年 3 月定例会
段落番号	2
役職名	議長
発言者名	櫻井正富
議員 ID	1
ファイルのパス	/甲府市議会/2010/平成 22 年 3 月定例会/平成 22 年 3 月定例会（第 1 号）.txt
発言	市長から、上程議案全部に対する提案理由の説明を求めます。
その他	

のコードが割り当てられているが、その他の委員会は市町村によって異なるためその他と一括りにしている。「年度」は表題に含まれる和暦を西暦に直している。「回」は表題に「第〇回定例会」のように書かれているものもあるが、例のように「〇月定例会」と書かれているものは、元ファイルが配置されていた同一ディレクトリ内の定例会の開催月を比較して何回目であるか

2,19201,0010,2010,1,3,定例会,1,0301,平成 22 年 3 月定例会,2,議長,櫻井正富,1,/ 甲府市議会/2010/
平成 22 年 3 月定例会/平成 22 年 3 月定例会 (第 1 号).txt, “市長から、上程議案前部に対す
る提案理由の説明を求めます。”

図 3 図 2 中の下線部に対して付与を行った結果

を推定している。「議会名」は表題から日付や回などを省くことで生成される。「号」はファイル名より会議が 1 日目であることを表している。「日付」はこの例の中には現れていないが、会議録の HTML タグの中に現れるものを抽出している。「表題」は会議録の HTML ファイルにある title 要素であるが、title がない場合はファイル名から「平成～年〇〇会」までを抽出している。「段落番号」は 2 番目の段落であることを表している。なお、段落の区切りは br タグにより判別される。「役職名」と「発言者名」は、発言者の発言の最初に、「〇役職名 (発言者名君)」のような表現で現れるものを抽出している。「議員 ID」は全国の地方議員の一覧を別途用意し、すべての議員に割り振った。「ファイルのパス」は収集したファイルの保存場所を示している。「発言」には、該当する 1 文の発言を文字列で保存している。「その他」には発言以外の会議録の内容、例えば「(市長 宮島雅展君 登壇)」のような記述が入れられる。

3.4 地方議会会議録コーパスを用いた研究

前節の手法により構築した地方議会会議録コーパス及びそのデータベースにより、全国の地方議会会議録に対し、「いつ」、「どの会議で」、「どの議員が」、「何を発言したのか」について検索を行うことが可能となる。これを受けて、今後、政治学、社会言語学、情報工学といった各分野での研究が期待されるが、以下では現時点で行われている、地方議会会議録コーパスを用いた情報工学と社会言語学の研究について紹介する。

3.4.1 情報工学の研究

情報工学の分野では、会議録に含まれるテキストから、政治的課題の表現や要求表現の自動抽出、抽出データの関係推定などを用いて、住民、自治体職員、政治家などに有益な情報を提供する研究が行われている。

会議録は定例会だけでも膨大な量であり、北海道小樽市の市議会会議録の場合、定例会 1 回分の会議録だけで A4 判にすると 200 ページを超える。木村ら (木村他 2011) は、大量のテキストデータに対して能動的にアクセスし、これらのデータを読む住民が少ないと考え政治的課題の関心を明確にするための質問をシステムから利用者に行うことで、利用者の考えに近い議員を提示する方法を提案している。

会議録に含まれる重要部分を抽出する研究も行われている。葦原ら(葦原, 木村, 荒木 2012)は, 会議録に含まれる重要な内容が議員からの質問に含まれることが多いことに着目し, 議員の質問から要求表現を抽出する研究を行っている。

他には, 大城ら(大城 他 2012)は施策や事業に対する賛否の意見を, 構造的にとらえるための注釈付けタグセットを提案している。議員の施策や事業の意見について注釈付けを行うという点では共通しているが, 同じ賛成(もしくは反対)を示す議員に対しその積極性を考慮するという点で本研究とは異なっている。

3.4.2 言語学の研究

地方議会会議録は, 社会言語学, 日本語学, 方言学などの研究に寄与する言語資源であると考えられる。しかし, 会議録は議会における発言を一字一句厳密に記録しているわけではなく, 文章としての読みやすさを考慮して, 意味内容が大きく変わらない範囲で修正(整文)が加えられている。高丸ら(高丸, 木村 2010; 高丸 2011, 2013)は地方議会会議録の言語資源としての性質を明らかにするための基礎研究として, 複数の地方議会会議録における整文の状況を分析し, 実態を比較した。

整文の過程において, 冗長な表現の削除や言い間違いや方言語彙の修正などが行われているため, 地方議会会議録コーパスを用い話し言葉などに含まれる非流暢性を分析することは困難であると考えられるが, 本コーパスは通時性・共時性を併せ持つ言語資源であるため, 新しい文法表現の需要の実態や議会用語の変遷等を分析することが可能である。さらに, 整文の担当者がある表現が方言であることに気付かないことや, 発言者の口調を維持するために, 方言であっても整文されずに残されることがあることといった理由により, 会議録に現れる方言等を分析することが可能である。現在これらの観点に基づく研究への本コーパスの活用が進められている。

4 賛否の積極性に関する注釈付け

3節で説明した地方議会会議録コーパスにより, 発言(文)や方言のような表層的な表現の検索や分析は可能となった。しかし, 我々が開発を目指している, 利用者の考えに近い議員を提示するシステムを構築するためには, ある議員の施策や事業に対する意見, 例えばその賛否や積極性を判定する必要がある。意見は複数の形態素等の要素を組み合わせることにより表されるものであると考えられるため, 先に述べたプレインテキストに基づくコーパスの構築のみでは不十分であると考え, 会議録中に表れる政治的課題や政策に対する賛否およびその積極性に関する注釈付けを行うこととした。

本節では, まず賛否の積極性に関する情報に関して考察し, 注釈すべき情報の定義を行う。付

与した XML 形式のタグの仕様と付与の基準について説明した後、注釈付け結果の統計を示す。最後に、タグに関する課題について述べる。

4.1 賛否の積極性に関する情報

賛否の積極性について考えるにあたり、会議録中の賛否を表明する発言を観察し、それらについて分析を行った。図 4 に、会議録の構成を 5 つの場面の観点から例文とともに示す。

図 4 に示した 5 つの場面の内、(i) から (iv) の場面において施策や事業およびそれに関する賛否を表す文や表現が現れることが多く、一方で (v) の場面ではほとんど現れなかった。

次に、議員が施策や事業に関する意見を述べる際の発言の例を図 5 に示す。施策や事業に関する意見を表す文は、例文 (1) の下線部 (a) のように施策や事業そのものを表す表現と、下線部 (b) のように発言者の施策や事業に関する意見の表現の 2 つから構成されることができ

- (i) 質問 (会派の代表としての代表質問、一議員としての一般質問の 2 種類がある)

例:私は、自民党横浜市議団を代表いたしまして、市政運営の重要課題について中田市長並びに田村教育長に質問いたします。(中略) そこで、開国博 Y 150 成功に向け重要な時期となる夏休みの集客策について市長の考えを伺います。

- (ii) 質問に対する回答

例:順次お答えを申し上げます。(中略) ヒルサイドエリアでの周辺地域の盛り上げ策でありまして、(中略) プレイメントの実施などを予定しております。

- (iii) 討論 (基本的に返答はない)

例:私は、日本共産党を代表して、今定例会に提出された市第 1 号議案横浜市職員に対する期末手当及び勤勉手当に関する条例の一部改正について反対の討論を行います。(中略) 今回の政治的で異常な勧告を認めるわけにはいきません。

- (iv) 議案、法案等の説明

例:市第 22 号議案について御説明いたします。本案は、一般会計補正予算 (第 2 号) でありまして、がん検診事業について国補正予算による国庫補助事業の詳細が示されたことから、本市における当該事業の迅速かつ円滑な実施を図るため、国庫支出金を財源として健康福祉費を 14 億 8675 万余円増額しようとするものであります。

- (v) 議長などによる挨拶、議決、各報告

例:ただいま書記に報告させましたとおり、現在着席議員数は 87 人であります。これより本日の会議を開きます。

図 4 会議録の場面構成

- (1) (a)生産年齢人口相の市民の定着が(b)必要と考えますが、将来に向けた本市の都市経営の在り方についてどのようにお考えなのかをお伺いします。
- (2) (c)トンネル工事の開始に向けて(d)積極的に取り組んでまいりたいと思います
- (3) 開港以来横浜に培われた
(e)国際性をさらに飛躍させていくという取り組みをこの近年
(f)強化しているわけでありまして、
- (4) 新型インフルエンザの国内発生後、直ちに発熱外来を9カ所開設することができまして、その後さらに4カ所の(g)増設を行いました。
- (5) 2007年からの団塊の世代の大量退職が始まり、時間的ゆとりを持つ高齢層の増加に加え、若年層についても余暇に対する意識の変化などにより、
(h)観光産業は今後の拡大が見込める成長産業と言えます。
(中略)
そこで、(i)国際観光コンベンション振興に関する市長の基本的な考え方についてお伺いいたします。
- (6) (j)一律に大幅な削減を行うような過度な見直しは(k)すべきでない
(l)要請してきたところ です
- (7) (m)発熱相談センターのオペレーターへの看護師の配置ということではありますが、発熱相談センターは、現在、オペレーターが相談シートに基づいて電話で相談を聞き取って、その内容をもとに医師、看護師を含む職員が判断を行うということにしているわけでもあります。職員は医療相談や発熱外来受診についての判断、医療機関との調整などを行っており、これは(n)適切に対応できていると思います。
- (8) (o)緊急借換支援資金の実施期間の延長についてではありますが、緊急借換支援資金は、現下の厳しい経済状況を考慮して、借入金の返済負担の軽減を図るために緊急対策として創設をしたものでありまして、(p)本年度末までの時限措置によって実施をしているものでございます。
- (9) 一方で競争を廃止した教育についても一部行われ、
(q)市民からも疑問の声が上がったと聞いております。

図 5 施策や事業に関する意見を含む文の例

る。なお、発言者としては、一般の議員に加えて、自治体の首長及び部局長、委員長、議長なども現れる。

まず、施策や事業の表現についての分析結果について述べる。例文(2)では、施策や事業は下線部(c)の複合名詞1語で表されているが、例文(1)の下線部(a)や例文(3)の下線部(e)、例

文(7)の下線部(m)のようにより大きな名詞句で1つの施策や事業が表現されているものも多く存在しており、いずれの場合も連続する文字列で現れていることが多かった。また、例文(4)のように、施策や事業は現れていないが、その実施の度合いのみを述べている文も多く存在していた。

次に、施策や事業に対する賛否の表現についての分析結果を述べる。例文(1)から例文(5)までのように、施策や事業に関して賛成を述べる表現と、例文(6)、(7)のような施策や事業に対して反対を述べる表現がある。その比率としては賛成が非常に多かった。これは自分の関心のある施策や事業と、行政側に実現させたい施策や事業について言及することが非常に多く見られたことによるものである。賛成を述べる際には、例文(1)の下線部(b)や、例文(5)の下線部(h)のように明確に賛成の意思を表す文が非常に多く見られた。下線部(d)のように「積極的に」などの言葉が入り、積極的な意思を示す表現も見られた。反対を述べる際には、賛成の場合と同様に例文(6)の下線部(k)と(l)のように明確に反対の意思を表明する文も存在するが、例文(7)のように、現状で十分であり、新たな行動を起こす必要がないため、下線部(m)の施策や事業には賛成ではないというように、消極的に反対を述べる文も存在していた。

また、例文(5)の下線部(h)と(i)のように、施策や事業を表す表現が意見を表明する文以降の文で発言され、それ以前の文にその施策や事業に対する説明があり、それが施策や事業に対する意見の理由となるような文も存在していた。他にも例文(9)のように、自らの意見ではなく市民の声として施策や事業に対して意見を述べる文も見られた。

さらに、同じように賛成の立場（もしくは、反対の立場）を表明していても、言及する施策や事業がどのくらい実現されているかの度合いにも差が見られた。その度合いは図6に示す4つに大別することができた。ここで、議員の施策や事業に関する意見を表明する発言を読み、そ

- 度合い1 例文(1)の下線部(b)や例文(2)の下線部(d)のようにまだ取り組んでおらず、意見を述べただけの状態。「～すべき」「～したい」、「～と思う、考える」等の意思を表す表現が主に見られた。
- 度合い2 例文(3)の下線部(f)のように実際に取り組み始めてはいるが、まだ終了していない状態。「～している」、「～してきた」のように進行相の表現が多く見られた。
- 度合い3 例文(4)の下線部(g)のように政治課題に取り組み、それがすでに終了している状態。「～した」のように過去時制の表現が多く見られた。
- 度合い4 例文(5)のように、政治課題に着手している、もしくはもう既に完了していて更に行動を拡大しようとしている状態。

図6 施策や事業の実現の度合い

の賛否への積極性を判断する場合を内省すると、積極性を判断するための手掛かりとして、少なくとも以下の3点があった。1点目は、発言中の「やむを得ない」などのほかに手立てがないことを示す表現である。この表現がある場合、消極的な賛成（もしくは、反対）であることが読み取れる。しかしながら、会議録の場合、一般的には積極性を示すはずの表現の存在が必ずしも字義通りの積極性を示すとは限らない。例えば、「取り組まねばならないと考えている」という表現の場合、一般的には積極的な賛成を示していると考えられるが、会議録における議員の発言においては、この表現だけをもって、積極的な賛成であると判断することは不適切である。なぜなら、会議録には特有の表現や言い回しがあり、議員は施策や事業に積極的に取り組むことが当然として捉えられているため、積極的であることを示す表現が常時の表現となっていることが多い。一方、消極的であることを示す表現に関しては、例えば「大幅な繰り入れについてはおのずから限界があると考えるので、値上げについては賛成とは言えないが、やむを得ないと考える」のように、現状が非常に厳しいため本心では実施したくないのだが現状を鑑みると実施せざるを得ないという意図で発言したと読み取れ、消極性が感じられる。すなわち、積極的であることを示す表現が常時の表現のため、表現が積極的なものに偏っており、積極的か消極的かを示す表現で、非対称性があるように思われる。

2点目は、言及された施策や事業の具体性である。積極的であるように読み取れる場合、その施策や事業への言及が具体的であることが多い。例えば、単に「取り組まねばならない」という発言よりも、「〇〇といった理由により△月までに取り組まねばならない」という発言の方が、積極的に取り組むという意思を読み取ることができる。具体性の有無を判断する手掛かりとして、実現したい施策や事業の詳細な内容や、現状を数値などを踏まえて言及するような構造等が考えられるが、本稿では施策や事業に対する意見の根拠となる理由に着目し、理由が述べられている文には具体性があると考えた。

3点目は、言及された施策や事業の実現度合いである。すなわち、これから取り組みたいという意思を表明しているだけなのか、それとも、既に施策や事業の一部に着手しているのか、といった違いにより、賛否の積極性を読み取ることができると考えられる。

施策や事業に対する賛否の積極性を判断する上で、上記の3点が手がかりになるということは仮説であるが、賛否の積極性に対する人間の判断結果と共に、これらの手がかりに関する表現の注釈付けを行うことにより、その仮説の分析、および、その分析結果に基づいて構築された積極性判断のための仕組みが正しく動作しうかどうかの確認が可能となる。

4.2 賛否の積極性に関する注釈付け

本稿では、地方議会会議録コーパスに収録されている、札幌市、横浜市、京都市、北九州市の4市の2010年の第2回定例会を対象に注釈付けを行った。この4市を対象とした理由は、政令指定都市であること、全国に散らばっていること、同一の記述形式の会議録を採用している

ことの3点による。注釈付けを行う単位として、「発話」、「段落」、「文」、「文字列」の4つの単位を用い、「文」、「段落」、「発話」を以下のように定義した。まず、4市の会議録の記述形式では、全ての文の最後が句点で終わっていることから、句点を「文」の境界とした。次に、同一の話題に関する文は1つの段落にまとめて記述されており、全ての段落の最初には空白が存在することから、行頭の空白を「段落」の境界とした。最後に、文の発言者に関する情報がコーパス中に収録されているため、発言者が同一人物である文の連続を1つの「発話」とした。

4.1節での議論を基に、表4に示す11種類のタグに関して注釈付けを行うこととした。先に述べた4つの都市に対して、8人の注釈者が1人2都市ずつ担当し注釈付けを行い、1都市につき4つのコーパスを作成することとした。注釈者の育った言語環境は全員日本語で、出身地は神奈川県が3人、静岡県が2人、愛知県が1人、岡山県が1人、佐賀県が1人であった。注釈作業にかかった時間は、おおむね1都市につき15時間から20時間程度であった。各注釈の説明を以下に述べる。

(1) 番目は、発言がどのようなシーンでなされたかの注釈であり、発話単位で付与する。議員が意見を述べることが多いシーンについて、発言が行われる場面ごとの比較や分析を行えるよう「質問」、「回答」、「討論」、「説明」の4シーンを想定し、各シーンを以下のように定義した。

- 「質問」：発言中に他者に対して回答を求める文が存在している。
- 「回答」：発言中に他者からの質問に対する回答となる文が存在している。
- 「討論」：自分の意見を一方的に表明している文が存在している。
- 「説明」：議案等の内容を説明する文が存在している。

上記の4シーンに当てはまらないシーンは「その他」として注釈づけを行った。シーンは排他的に注釈付けられる。すなわち、注釈者は上記の4シーンに「その他」を加えた5つのうちか

表 4 注釈の一覧

番号	内容	付与単位
(1)	発言されたシーン	発話
(2)	施策・事業	文字列
(3)	発言内容のカテゴリー	段落
(4)	質問と回答の対応付け	段落→段落
(5)	疑問文	文
(6)	意見性の有無	文
(7)	意見の本人性	文
(8)	発言内容の中核となる文	文
(9)	理由の対応付け	文→文（文字列）
(10)	意見の実現度	文
(11)	意見の説得性	文

ら1つを選ぶ。作業効率の観点から、「その他」のシーンにおける発言には(2)以降の注釈付けを行わなかった。

(2) 番目は、発言者の関心がある施策や事業に関する注釈であり、文字列単位で付与する。発言中に含まれる施策や事業を示す文字列を同定するとともに、これに対し、「賛成(推進)」、「反対(廃止)」、「その他」の何れかの極性を付与する。これにより、発言に対する賛否の自動判定を行うための機械学習の教師情報として、注釈付けを行ったコーパスを利用できる。極性の判断は前後の文脈に現れる記述により、作業者の主観に基づいて行われた。施策や事業に関する注釈付けは次のような形で行われる。

<Policy Polarity="賛成">A N A 5 路線の存続</Policy>

この例では、「A N A 5 路線の存続」という施策に対し、発言者は賛成の意思を立場を示している。

(3) 番目は、発言内容のカテゴリに関する注釈であり、段落単位で付与する。カテゴリは、木村ら(木村, 渋谷, 高丸, 小林, 森 2010)の政治的カテゴリを参考に、比較的議題に挙げられることが多い、「医療」、「教育」、「環境」、「観光」、「防災」、「公共」の6カテゴリを対象とした。1つの段落に複数のカテゴリを付与することを許可している。また、発言内容がどのカテゴリにも属さない場合には「その他」として注釈付けを行った。

(4) 番目は、質問と回答の対応付けに関する注釈であり、段落単位で付与する。質問の段落から回答の段落へと1対1で対応付けており、もしも、回答が複数の段落にまたがっている場合は最初の段落に対応付けを行った。段落に対する注釈付けは次のような形で行われる。これにより、ある議員の質問とそれに対する行政側の回答が結び付き、施策・事業ごとの議員の意見と行政側の意見を1つの組として分析が可能となる。

<Paragraph Id="P338" CorrespondingAnswerParagraphID="P438" Category="医療">

この例では、338番目の段落は医療のカテゴリについて発言しており、その段落の中で現れた質問は、438段落で回答されていることを表している。

(5) 番目は、疑問文かどうかを判断した結果の注釈であり、文単位で付与する。本稿での疑問文とは、他者の回答を要求する文と定義しており、「〇〇についてお聞かせ願いたい」といった表現であっても疑問文とした。

(6) 番目は、意見性がある文かどうかを判断した結果の注釈であり、文単位で付与する。本稿での意見性がある文とは、「〇〇すべきだ」、「△△の方が良いと考えられる」といった意見であることが明確に示されている文と定義している。

(7) 番目は、発言者本人の意見である文かどうかを判断した結果の注釈であり、文単位で付与する。

(8) 番目は、発言内容の中核となる文に関する注釈である。本稿での中核となる文とは、発言内容を端的に述べている文と定義している。我々は、システムが利用者に発言内容を提示する

際には、発言内容の整理・要約を行う必要があると考えており、整理・要約を行うための情報として利用することを想定している。(2)の施策・事業を注釈を含む文、または、(6)の意見性があると判断された文を含む段落には、最低でも1文は中核となる文を選定し注釈付けを行うこととした。

(9) 番目は、(2)の施策・事業の極性または(6)の意見性がある文と、その理由となる文との対応付けに関する注釈である。理由となる文から、(2)で選択された施策・事業または(6)の意見性がある文へと1対多で対応付けを行っている。これらが複数存在することにより、理由となる文の集合と、(2)ならびに(6)に属する文の集合の間に多対多の関係が成り立つ。これにより、施策や事業に対する賛否の現れ方や、どのような発言が賛否の理由となるのかについての分析が可能となる。

(10) 番目は、発言時点で文中の意見がどの程度実現されているかの注釈であり、文単位で付与する。実現の程度として、「表明」、「着手」、「完了」、「拡大」の4つの状態を以下のように定義した。

- 「表明」：何も実現できていない状態。やるべきという意味を表明しただけの状態。
- 「着手」：実現のために行動を開始した状態。現在進行中であり、目標は達成されていない。
- 「完了」：すでに目標を達成した状態。現在は行動していない。
- 「拡大」：すでに目標を達成しており、さらなる成果を求めて行動したい(している)状態。

(2)の施策・事業を注釈した文には必ず付与することとした。

(11) 番目は、総合的に見て、文中の意見がどの程度説得性がありそうか（目標を実現できそうか）に関する注釈であり、文単位で付与する。説得性の判断は前後の文脈を考慮した作業者の主観に基づいて行われ、(2)の施策・事業に関する注釈付けを行った文には必ず付与することとした。前述の理由の対応付けや意見性の有無と合わせて、発言中のどの要素が意見の積極性を表すかについての考察が可能となる。文に関する注釈付けは以下のように行われる。

<Sentence Id="S346" Member="(小川直人議員)" IsQuestion="False" IsOpinion="True" IsPrincipal="True" Actualization="不明" CorrespondingConclusive="P:352_11_9: A N A 5 路線の存続" IsPersuasive="True" IsCoreSentence="False">丘珠5路線は、道内主要都市を結び、ビジネスマンや観光客、さらには札幌市内医療機関への通院など、多くの人がさまざまな目的を持ち札幌と各地を往来しており、移転することで年間37万人の利用者の利便性や経済活動を著しく損なうことになるのは明らかであります。</Sentence>

<Sentence Id="S352" Member="(小川直人議員)" IsQuestion="True" IsOpinion="True" IsPrincipal="True" Actualization="表明" CorrespondingConclusive="None" IsPersuasive="True" IsCoreSentence="True">そこで、質問ですが、<Policy Polarity="賛成">A N A 5 路線の存続</Policy>に向けては、道、経済界、関係自治体が一体となった活動が求めら

政治情報コーパス作成ツール -yokohama_2010_2.csv- [F1:読込] [F2:保存] [F5:XML出力] [F9:チェック] [F12:印刷...]									
発言者	Q→A	理→結	発言<434文/3766文>	疑問	意見	本人	実現度		
(中山大輔君)	○	1/1 ●	他の医療機関に協力を依頼することにより一時的に入院患者への対応はできると思いますが、現行では厳重な感染対策を講じ、さらに、人工呼吸器の装着が必要な重症患者が増加をすれば通常の診療に影響が出ることなど、強毒性のインフルエンザの入院治療必要患者を受け入れられる余力が一般の病院にあるのか、懸念されるところです。	×	○	○	表明	着手	完了 拡大
		1/1 ●	現段階のように、新型インフルエンザは毎年流行するインフルエンザと同じ弱毒性との見解を示されていますが、しかしながら、世界じゅうの大半の人が経験をしたことがないウィルスのため、免疫がなく、さらに、インフルエンザウィルスは変異をしやすく、流行中に強毒性に変わることも考えられるため、警戒は依然怠れないと考えます。	×	○	○	表明	着手	完了 拡大
		1/1 ●	そうした中、パンデミックの状態になった場合に備え、感染症病床を方面別に新たに整備するなどの対策も必要ではないかと考えます。	×	○	○	表明	着手	完了 拡大
	●	○	そこで、感染症病床確保について方面別整備の考え方を取り入れ、地域中核病院等を活用しながら一定数以上の感染症病床を備えた拠点的な施設を整備し、い	○	○	○	表明	着手	完了 拡大

図 7 タグ付けツール

れ、さらには、道民、市民に大きく運動を広げていくことも視野に入れた取り組みが必要と考えますが、今後、市長はどのように対応しようとしているのか、お伺いいたします。

</Sentence>

表 4 中の各注釈は (5) から順に、「IsQuestion」、「IsOpinion」、「IsPrincipal」、「IsCoreSentence」、「CorrespondingConclusive」、「Actualization」、「IsPersuasive」と表され、この例では、2 文目に現れている「ANA 5 路線の存続」という施策に対し、1 文目がその理由として結びついている。

注釈者の注釈付けの際の誤りを軽減するために、図 7 に示す、専用のタグ付けツールを開発し、ツールを通して上記の注釈付けを行った。注釈情報は、図 8 に示すような XML 形式で付与される。

4.3 注釈結果の統計および分析

4.2 節で述べた 4 都市の当該会議録に対する注釈付けを行った結果の傾向を分析するために、各統計量を調査した。その結果を表 5 から表 10 に示す。以下の (1) などの数字は、4.1 節の表 4

```

<Paragraph Id="P178" CorrespondingAnswerParagraphID="None" Category="医療">
<Sentence Id="S432" Member="(中山大輔君)" IsQuestion="False" IsOpinion="True"
IsPrincipal="True" Actualization="不明" CorrespondingConclusive="P:434_25_17:感染症病床を
方面別に新たに整備する" IsPersuasive="False" IsCoreSentence="True"> 他の医療機関に協
力を依頼することにより一時的に入院患者への対応はできると思いますが、現行では 심각한
感染対策を講じ、さらに、人工呼吸器の装着が必要な重症患者が増加をすれば通常の診療に
影響が出ることなど、強毒性のインフルエンザの入院治療必要患者を受け入れられる余力が
一般の病院にあるのか、懸念されるところです。</Sentence>
<Sentence Id="S433" Member="(中山大輔君)" IsQuestion="False" IsOpinion="True"
IsPrincipal="True" Actualization="不明" CorrespondingConclusive="P:434_25_17:感染症病床を
方面別に新たに整備する" IsPersuasive="False" IsCoreSentence="False">現段階のように、新型
インフルエンザは毎年流行するインフルエンザと同じ弱毒性との見解を示されていますが、し
かしながら、世界じゅうの大半の人が経験をしたことがないウイルスのため、免疫がなく、さら
に、インフルエンザウイルスは変異をしやすく、流行中に強毒性に変わることも考えられるため、
警戒は依然怠れないと考えます。</Sentence>
<Sentence Id="S434" Member="(中山大輔君)" IsQuestion="False" IsOpinion="True"
IsPrincipal="True" Actualization="表明" CorrespondingConclusive="P:434_25_17:感染症病床を
方面別に新たに整備する" IsPersuasive="True" IsCoreSentence="True">そうした中、パンデミッ
クの状態になった場合に備え、<Policy Polarity="賛成">感染症病床を方面別に新たに整備す
る</Policy>などの対策も必要ではないかと考えます。</Sentence>
</Paragraph>

```

図 8 XML データの例

表 5 総発話数とシーンごとの内訳

	横浜市	札幌市	北九州市	京都市
発話数	155	151	343	110
質問	22 (.142)	11 (.073)	33 (.096)	9 (.082)
回答	32 (.206)	32 (.212)	120 (.350)	27 (.245)
討論	6 (.039)	2 (.013)	5 (.015)	2 (.018)
説明	6 (.039)	3 (.020)	6 (.017)	4 (.036)
その他	89 (.574)	103 (.682)	179 (.521)	68 (.618)

表 6 段落とカテゴリーごとの内訳

	横浜市	札幌市	北九州市	京都市
段落数	1,618	1,797	2,385	474
カテゴリー	1,222	1,090	2,050	249
医療	243 (.199)	242 (.222)	528 (.258)	70 (.281)
教育	195 (.159)	151 (.139)	304 (.148)	46 (.185)
環境	75 (.061)	80 (.073)	224 (.109)	15 (.060)
観光	66 (.054)	82 (.075)	109 (.053)	2 (.008)
防災	39 (.032)	19 (.017)	48 (.023)	0 (.000)
公共	192 (.157)	186 (.170)	389 (.190)	44 (.177)
その他	553 (.453)	397 (.364)	756 (.369)	92 (.369)

表 7 文単位の注釈結果

	横浜市	札幌市	北九州市	京都市
文数	3,766	2,480	4,665	1,827
質問	1,677 (.445)	1,152 (.465)	2,176 (.466)	865 (.473)
回答	1,075 (.285)	549 (.221)	1,924 (.412)	374 (.205)
討論	328 (.087)	97 (.039)	41 (.009)	37 (.020)
説明	40 (.011)	65 (.026)	101 (.022)	32 (.018)
その他	646 (.172)	617 (.249)	423 (.091)	519 (.284)
カテゴリー	3,120	1,863	4,242	1,308
医療	575 (.184)	395 (.212)	1,171 (.276)	491 (.375)
教育	420 (.135)	239 (.128)	592 (.140)	299 (.229)
環境	172 (.055)	136 (.074)	450 (.106)	94 (.072)
観光	195 (.006)	150 (.081)	232 (.055)	13 (.010)
防災	82 (.026)	33 (.018)	90 (.021)	0 (.000)
公共	192 (.157)	336 (.180)	792 (.187)	268 (.205)
その他	1,557 (.499)	710 (.381)	1,552 (.366)	286 (.219)
疑問文	277	160	312	99
意見文	951	524	736	221
本人性	916	496	713	212
中核文	885	605	990	296
理由文	719	534	518	194
実現度	688	467	510	254
表明	466 (.677)	318 (.681)	339 (.665)	151 (.594)
着手	165 (.240)	117 (.250)	133 (.260)	74 (.291)
完了	13 (.019)	5 (.011)	13 (.025)	29 (.114)
拡大	44 (.064)	27 (.058)	25 (.049)	0 (.000)
説得性	413	225	205	141

表 8 施策・事業の注釈結果

	横浜市	札幌市	北九州市	京都市
施策・事業数	424	336	484	142
賛成	387 (.913)	302 (.899)	374 (.773)	117 (.824)
反対	35 (.082)	30 (.090)	91 (.188)	20 (.141)
その他	2 (.005)	4 (.012)	19 (.039)	5 (.035)

表 9 実現度と説得性の内訳

	横浜市	札幌市	北九州市	京都市
説得性	413	225	205	141
実現度の注釈がつけられている文	688	467	510	254
実現度の注釈がつけられている文で説得性のあるもの	335 (.508)	194 (.415)	194 (.380)	97 (.382)
実現度の注釈がつけられていない文	2,432	1,396	3,732	1,054
実現度の注釈がつけられていない文で説得性のあるもの	78 (.032)	31 (.022)	11 (.003)	44 (.041)

表 10 理由と説得性の関係

	説得性あり	説得性なし	計
理由を持つ文	627	475	1,102
理由を持たない文	357	1,628	1,985
計	984	2,103	3,087

中の注釈の種類番号である。表 5 に、会議録中の発話の数と (1) で付与されたシーンの内訳を示す。表 6 には、会議録中の段落の数、(3) のカテゴリーが付与された段落の数およびその内訳を示す。各段落には複数のカテゴリーの付与を許可していることと、ならびに「その他」のシーンの段落にはカテゴリーが付与されていないことに注意されたい。表 7 に、会議録中の文の数、(1) で付与されたシーンの文の数とその内訳、(3) で付与されたカテゴリーの文の数とその内訳、(5) の疑問文の数、(6) の意見性のある文の数、(7) の本人の意見である文の数、(8) の中核となる文の数、(9) の理由となる文の数、(10) の実現度を有する文の数とその内訳、(11) の説得性がある文の数を示す。表 8 には、(2) の抽出された施策や事業及び極性の内訳を示す。各表中の括弧内の値は割合を示している。表 9 に、各実現度と説得性の関係を示しており、括弧内の数はそれぞれ表明の文の中で説得性のあるものの割合と、着手・完了・拡大の文の中で説得性のあるものの割合を示している。表 10 に、意見文および施策や事業に対して理由が結びつくかどうか、それらが説得性を持つかどうかを示している。なお、表中の数値は 4 都市の注釈結果を合計したものである。

まず、4 都市間で注釈結果を比較すると、表 5 から表 8 に関しては注釈の数の分布に大きな差は見られない。

次に、表 5 から順に統計量からわかったことについて述べる。

表 5 の総発話数とシーンごとの内訳を見ると、いずれの市においても「その他」のシーンが一番多く、「回答」、「質問」のシーンが残りの大部分を占めている。「その他」のシーンは、図 4 の (5) に示したとおり、議長の挨拶や議決、予算などの各種報告に対して注釈されるものであり、施策や事業に対し意見を述べる発言はほぼ存在しない。そのため、施策や事業に対する賛否を判定する際には不要であり、シーンに関する注釈付けをおこなうことにより、これらを省いたデータを作成することが可能になると考えられる。また、「回答」のシーンの注釈数を 4 都市間で比較すると、北九州市のみ他の都市より多くなっていた。4 都市とも質問者の多数の質問に対し、市長及び関係部署の議員が答える形式をとっているが、北九州市は部署の区切りが「建設都市局長」「建設局長」というように役職が細かく設定されているため、回答者が増える傾向にあるからではないかと思われる。このように、各シーンの分布の違いから各都市の議会の傾向について分析することも可能であることがわかった。

表 6 の段落とカテゴリーごとの内訳を見ると、いずれの都市においても「その他」が一番多

く、「医療」が次に続き「教育」、「公共」が残りの多くを占めている。今回注釈付けを行ったのは2010年の会議録であり、その前年の2009年に新型インフルエンザが世界的に流行していたため、それに対しての対策等について述べる議員が多く、「医療」のカテゴリーであると注釈された段落の数が他の物に比べて多くなったものと考えられる。この結果から、議会の話題は時事的な問題に影響を受け得ることがわかった。

表7の文単位の注釈結果を見ると、カテゴリーの分布は特に変わらないが、シーンの分布は大きく変化し、「質問」のシーンと注釈された発話に含まれる文が一番多くなっている。これは4都市の議会の質問応答形式が一括質疑・一括答弁であり、発話者が交代する場面が少ないことによる。このように、発話単位でのシーンと各シーンの発話に含まれる文の数を比較することで、議会の質問形式の傾向を知ることができる。

表8の施策や事業の極性の内訳をみると、賛成の割合が横浜市、札幌市においては約9割、北九州市、京都市においては約8割をそれぞれ占めていた。これにより、会議録中の発言では賛成意見が常時の表現となっているという4.1節における1点目の観察について裏付けることができた。また、賛否の判定を二値分類で行う際には否定の判定精度が重要になると考えられる。

表9の実現度と説得性の関係を見ると、実現度に関する注釈付けのある文に関して説得性があると判断された文の割合が約4割から5割となっており、実現度に関する注釈付けのない文に関して説得性があると判断された文の割合を大きく上回っている。これにより、度合いにかかわらず実現度を含む文、すなわち施策や事業を実現したいという発言や、既に実施しているという発言は積極性に寄与すると考えることができそうである。

最後に、意見文および施策や事業に対して、それらに理由が結びついているか、説得性を持つかに関しての統計を示した表10を見ると、理由が結びついている文や施策は説得性があると判断されることが多い傾向にあり、理由を持つ文は積極性の手がかりになるという4.1節における2点目の観察を裏付けることができた。

4.4 付与したタグの課題

本小節では地方議会会議録コーパスに我々が提案したタグを付与した際に明らかになった課題を、「タグの仕様」と「注釈付けにおける主観的判断」の2つの観点から説明する。

4.4.1 タグの仕様策定に関わる課題

「理由の対応付け」のタグは施策や事業に関する賛否の理由となる文に対して付与され、その理由と施策や事業とを結びつけるものであるが、その付与単位を「文」としている。また、会議録では「〇〇してまいります」や「〇〇と考えます」等のような文末表現が常態となっており、「××を行うことは大きな利益を生むため、推進していこうと考えております」といった、理由を含む複文においても文末表現として現れることが多い。そのため、理由を表す表現につ

いて分析を行いたい時には、単に理由を含むと注釈づけられた文全体に注目するだけでは正しくないという問題がある。理由に纏わる分析を簡単に行えるようにするためには、「理由の対応付け」のタグの注釈単位を文字列とし、先述の例の「××を行うことは大きな利益を生むため」にのような従属節等に注釈付けを行えるような仕様とすれば、この問題は解決できると考えられる。

4.4.2 注釈付けにおける主観的判断

テキストに対する注釈付けにおいて、一般的に、注釈者間の判断が必ずしも一致しないことが問題となる。本稿における注釈付けにおいては、特に「施策・事業」の認定に関する判断に揺れが見られた。「施策・事業」のタグは文字列に対し付与されるものであるが、議員の発言中に必ずしも施策や事業がひとつの連続した文字列や名詞句の形で出現するわけではなく、図9のように、注釈者によりどの範囲を施策や事業として捉えるのかが異なってしまうことがあった。いずれの例もどの範囲を施策や事業として注釈付けるかで揺れが生じているのだが、下の例に関してはある注釈者は2つの施策があると注釈し、別の注釈者は纏めて1つの施策として注釈付けている。これにより施策や事業とそれに関する賛否の分析を行う際にばらつきが生じてしまう。解決策としては施策や事業のみ範囲をあらかじめ決めておくという仕様にする可以考虑。

- (中略) 各商店街が行うキャンペーン事業や広告宣伝事業等を支援するために要する経費を追加するものであります。
- (中略) 各商店街が行うキャンペーン事業や広告宣伝事業等を支援するために要する経費を追加するものであります。
- (中略) 消費者センターの消費生活相談窓口や消費者への教育、啓発の強化等の事業を実施するための経費を計上するとともに、消防法施行令の改正により、スプリンクラーの整備が必要となる認知症高齢者グループホームに対し、補助を行うための経費を計上するものであります。
- (中略) 消費者センターの消費生活相談窓口や消費者への教育、啓発の強化等の事業を実施するための経費を計上するとともに、消防法施行令の改正により、スプリンクラーの整備が必要となる認知症高齢者グループホームに対し、補助を行うための経費を計上するものであります。

図9 施策や事業の認定に関する判断に生じた揺れの例

4.4.3 提案した注釈付け手法に追加して必要であると考えられる情報

本稿では議員の施策や事業に関する賛否と積極性に関する注釈付け手法を提案したが、積極性は必ずしもその有無といった二値の値で判断される情報ではなく、例えば「やや積極的」「かなり積極的」というように、積極的な場合と消極的な場合のいずれにおいても複数の段階が存在することがある。このような場合において、より詳細に注釈付けを行うためには、積極性に関する度合いを表現する手段が追加される必要がある。

5 おわりに

本稿では、地方政治に関する研究の活性化・学際的応用を目指して、全都道府県の県庁所在地および政令指定都市の計 51 市町村について会議録の収集とコーパスの構築を行った。51 市町村の会議録は Web 上で主に 4 社の会議録検索システムにより提供されており、ページに張られたリンクをたどっていく方法と CGI のパラメータを変えていく方法などにより、会議録を自動的に収集することを行った。地方議会会議録コーパスには 17 項目の情報を付与しており、また、発言単位に分割してデータベース化を行っている。

また、我々が目的とする会議録における発言を基に利用者と政治的に近い考えをもつ議員を判断して提示するシステムの開発に向け、地方議会会議録コーパスの分析・評価用のデータ作成のために、会議録中の議員の、施策や事業に対する賛否とその積極性に関する注釈付けを行う手法について提案をした。注釈結果の統計および課題について論じた。今後、本稿での分析により得られた知見を基に、各発言に対して、賛否とその積極性を自動判定する手法を開発したいと考えている。

謝 辞

本研究の一部は、JSPS 科研費 22300086 の助成を受けたものである。

参考文献

- 葦原史敏, 木村泰知, 荒木健治 (2012). 地方議会会議録における要求・要望表現抽出の提案. 言語処理学会第 18 回年次大会論文集, pp. 1-27.
- de Goede, B., van Wees, J., Marx, M., and Reinanda, R. (2013). “Political Mashup Ngramviewer Tracking Who Said What and When in Parliament.” *Research and Advanced Technology for Digital Libraries*, pp. 446-449.

- 筒井, 我満, 大城, 菅原, 永井, 渋谷, 木村, 森 地方議会会議録コーパスと政治情報システム構築のためのアノテーション
- 橋本力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明 (2011). 構文・照応・評価情報つきブログコーパスの構築. 自然言語処理, **18** (2), pp. 175–201.
- 平田佐智子, 中村聡史, 小松孝徳, 秋田喜美 (2012). 国会会議録コーパスを用いたオノマトペ使用の地域比較. 第 27 回人工知能学会全国大会論文集, 3N4-OS-01c-2, **31** (10).
- 飯田龍, 小町守, 乾健太郎, 松本裕治 (2008). 述語項構造と照応関係のアノテーション: NAIST テキストコーパス構築の経験から. 自然言語処理, **17** (2), pp. 25–50.
- 井上史雄 (2013). [ことばの散歩道] 171 去った〇日. 日本語学, **31-10**.
- 川端正法, 山本和英 (2007). 話題の継続に着目した国会会議録要約. 言語処理学会第 13 回年度大会, pp. 696–699.
- 川浦昭彦 (2009). “Self-Serving Mayors and Local Government Consolidations in Hokkaido.” 日本経済学会春季大会研究報告.
- Kawaura, A. (2010). “Self-Serving Mayors and Local Government Consolidations in Japan.” In *University of Hawaii Department of Economics Working Paper*.
- 木村泰知, 渋谷英潔, 高丸圭一 (2009). 地方議員と住民間の共同支援に向けたウェブの利用. 選挙研究, **25** (1), pp. 110–118.
- 木村泰知, 渋谷英潔, 高丸圭一, 小林哲郎, 森辰則 (2010). 北海道を対象とした地方議員と住民間の共同支援システムのユーザインターフェース評価. 第 24 回人工知能学会全国大会論文集, 2J2-NFC2-3.
- 木村泰知, 渋谷英潔, 高丸圭一, 乙武北斗, 小林哲郎, 森辰則 (2011). 地方議員マッチングシステムにおける能動的質問のための質問生成手法. 第 24 回人工知能学会全国大会論文集, **26** (5), pp. 580–593.
- 小林のぞみ, 乾健太郎, 松本裕治 (2006). 意見情報の抽出／構造化のタスク使用に関する考察. 自然言語処理研究会報告 2006-NL-171.
- 松吉俊, 江口萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治 (2010). テキスト情報分析のための判断情報アノテーション. 電子情報通信学会論文誌, **J93-D** (6), pp. 705–713.
- 宮崎林太郎, 森辰則 (2010). 注釈事例参照を用いた複数注釈者による評判情報コーパスの作成. 自然言語処理, **17** (5), pp. 3–50.
- 森脇俊雅 (2008). 合併と地方議会活動: 議員アンケートの分析を中心にして. 選挙研究, **23**, pp. 82–90.
- 大城卓, 渡邊裕斗, 渋谷英潔, 木村泰知, 森辰則 (2012). 地方政治情報システムのための地方議会会議録への注釈付けタグセットの提案. 言語処理学会第 18 回年次大会発表論文集, pp. 3–9.
- 乙武北斗, 高丸圭一, 渋谷英潔, 木村泰知, 荒木健治 (2009). 地方議会会議録の自動収集に向けた公開パタンの分析. 言語処理学会第 15 回年次大会, pp. 192–195.
- Ptaszynski, M., Rzepka, R., Araki, K., and Momouchi, Y. (2012). “Automatically Annotating A

- Five-Billion-Word Corpus of Japanese Blogs for Affect and Sentiment Analysis.” In *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pp. 123–130.
- 関口洋一, 山本和英 (2003). Web コーパスの提案. 情報処理学会研究報告. 情報学基礎研究会報告, **2003** (98), pp. 123–130.
- 高丸圭一, 木村泰知 (2010). 栃木県の地方議会会議録における整文についての基礎分析—本会議のウェブ配信と会議録の比較—. 都市経済研究年報, pp. 74–86.
- 高丸圭一 (2011). 規模の異なる自治体における地方議会会議録の整文の比較. 社会言語科学会第 27 回研究大会, pp. 256–259.
- 高丸圭一 (2013). 形態素 N-gram を用いた地方議会会議録における地域差の分析手法の検討—ひらがなで構成された文末の 4-gram に着目して—. 明海日本語, pp. 1–10.
- 上神貴佳 (2006). 投票支援ツールと『政策中心の選挙』の実現—オランダの実践と日本における展望—. 選挙学会紀要, **6**, pp. 43–64.
- 上神貴佳, 堤英敬 (2008). 投票支援のためのインターネット・ツール—日本版ボートマッチの作成プロセスについて—. 選挙学会紀要, **10**, pp. 39–80.
- 上神貴佳, 佐藤哲也 (2009). 政党や政治家の政策的な立場を推定する—コンピュータによる自動コーディングの試み—. 選挙研究, **25** (1), pp. 61–73.
- Weibe, J., Wilson, T., and Cardie, C. (2005). “Annotating Expressions of Opinions and Emotions in Language.” *Language Resources and Evaluation*, **39** (2–3), pp. 165–210.
- 山本和英, 安達康昭 (2005). 国会会議録を対象とする話し言葉要約. 自然言語処理, **12** (1), pp. 51–78.
- Yu, B., Kaufmann, S., and Diermeier, D. (2008). “Classifying Party Affiliation from Political Speech.” *Journal of Information Technology & Politics*, **5** (1), pp. 33–48.
- 平野淳一 (2008). 「平成の大合併」と市長選挙. 選挙研究, **24** (1), pp. 32–39.

付録

A 各会議録検索システムのクロールプログラムの仕様

表 1 に示した 4 つの会社が提供する会議録検索システムのそれぞれに収録された会議録を収集するためのクロールプログラムの仕様は以下のとおりである。

- 大和速記情報センター・会議録研究所：大和速記情報センターおよび会議録研究所の会議録検索システムを導入している市町村の Web ページでは、トップページもしくはトッ

ページから直接リンクが張られているページに、各年度の会議録へのリンク一覧が存在するものがほとんどである。リンク一覧が存在しない場合、検索用の入力フォームに未記入で検索を実行することで全会議録の検索結果がリンク情報として表示される。これらのリンクをクロールプログラムが辿ることで会議録のページを自動的に取得する。

- フューチャーイン：フューチャーインの会議録検索システムを導入している市町村の Web ページでは、会議録検索システムの出力が CGI プログラムにより自動生成されていて、その CGI プログラムに渡すパラメタにより出力内容を制御できる。例えば、会議録の検索は、以下のようなパラメタを渡すことで行われている。

```
ACT=100&KENSAKU=0&SORT=0&KTYP=0,1,2,3,4&KGTP=0,1,2,3,4&PAGE=1
```

CGI プログラムに渡すパラメタ PAGE の値を順次変えることですべての検索結果を得ることができる。パラメタ ACT, KTYP の値はそれぞれ、ページの表示方法、会議種別に対応する。また、会議録は発言ごとに分割されており、同じ CGI プログラムにおいて、次のようなパラメタを渡すことで各発言を取得できる。

```
ACT=203&KENSAKU=0&SORT=0&KTYP=2,3&KGTP=1,2&TITL.SUBT=%95%BD%90%AC%82Q%82Q%94N%81@%82Q%8C%8E%92%E8%97%E1%89%EF%81%7C03%8C%8E03%93%FA-04%8D%86&HUID=46845&FINO=655&HATUGENMODE=0&HYOJIMODE=0&STYLE=0
```

パラメタ TITL.SUBT, HUID の値はそれぞれ、URI エンコードされた表題、発言 ID に対応する。パラメタ ACT は、この例の「203」では「発言」、ひとつ前の例の「100」では「検索結果」を指している。パラメタ TITL.SUBT の値は、この例では「平成 22 年 2 月定例会 03 月 03 日-04 号」を指している。クロールプログラムは、これらパラメタの値を順次変えることで CGI プログラムを経由して会議録を自動的に取得する。

- 神戸総合速記：神戸総合速記の会議録検索システムを導入している市町村の Web ページでは、会議録検索システムの検索結果の出力が CGI プログラムにより自動生成されており、その CGI プログラムに渡すパラメタにより出力内容を制御できる。例えば、会議録の検索は、CGI プログラムに以下のようなパラメタを渡すことで行われている。

```
treedepth=%95%BD%90%AC22%94N%20%95%BD%90%AC22%94N%203%8C%8E%92%E8%97%E1%89%EF%20
```

パラメタ treedepth の値は URI エンコードされた和暦と表題に対応する。この例では「平成 22 年 3 月定例会」を指している。クロールプログラムは、このパラメタの値を変え、CGI プログラムが生成したページに張られたリンクをたどることで会議録のページを自

動的に取得する。

略歴

筒井 貴士：2013 年横浜国立大学工学部電子情報工学科卒業。現在，同大学大学院環境情報学府情報メディア環境学専攻博士課程前期在学中。自然言語処理に関する研究に従事。

我満 拓弥：2013 年横浜国立大学工学部電子情報工学科卒業。現在，東京大学大学院工学系研究科電気系工学専攻博士課程前期在学中。数理生命情報学に関する研究に従事。

大城 卓：2010 年横浜国立大学工学部電子情報工学科卒業。2012 年同大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了。修士（情報学）。在学中は自然言語処理に関する研究に従事。

菅原 晃平：2010 年横浜国立大学工学部電子情報工学科卒業。2012 年同大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了。修士（情報学）。在学中は自然言語処理に関する研究に従事。

永井 隆広：2010 年横浜国立大学工学部電子情報工学科卒業。2012 年同大学大学院環境情報学府情報メディア環境学専攻博士課程前期修了。修士（情報学）。在学中は自然言語処理に関する研究に従事。

渋谷 英潔：1997 年小樽商科大学商学部商業教員養成課程卒業。1999 年同大学大学院商学研究科修士課程修了。2002 年北海道大学大学院工学研究科博士後期課程修了。博士（工学）。2006 年北海学園大学大学院経営学研究科博士後期課程修了。博士（経営学）。現在，横浜国立大学環境情報研究院科学研究費研究員。自然言語に関する研究に従事。言語処理学会，情報処理学会，電子情報通信学会，日本認知科学学会各会員。

木村 泰知：2004 年北海道大学大学院工学研究科電子情報工学専攻博士後期課程修了。博士（工学）。2005 年，小樽商科大学商学部助教授着任。2007 年，同准教授，現在に至る。この間，2010 年 10 月より 2011 年 9 月まで New York 大学客員研究員。自然言語処理，情報抽出などの研究に従事。言語処理学会，人工知能学会，情報処理学会，電子情報通信学会各会員。

森 辰則：1991 年横浜国立大学大学院工学研究科博士課程後期修了。工学博士。同年，同大学工学部助手着任。同講師，同助教授を経て，現在，同大学大学院環境情報研究院教授。この間，1998 年 2 月より 11 月まで Stanford 大学 CSLI 客員研究員。自然言語処理，情報抽出，情報検索などの研究に従事。言語処理学会，人工知能学会，情報処理学会，電子情報通信学会，ACM 各会員。

筒井, 我満, 大城, 菅原, 永井, 洪木, 木村, 森 地方議会議録コーパスと政治情報システム構築のためのアノテーション

(2013 年 9 月 20 日 受付)

(2013 年 12 月 2 日 再受付)

(2014 年 1 月 10 日 採録)